(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 October 2000 (26.10.2000)

PCT

(10) International Publication Number
WO 00/63382 A1

(51) International Patent Classification[7]: C12N 15/12, C07K 14/52, G01N 33/50, C07K 16/24

(21) International Application Number: PCT/US00/09904

(22) International Filing Date: 14 April 2000 (14.04.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/129,463 15 April 1999 (15.04.1999) US

(71) Applicant *(for all designated States except US)*: **OSIRIS THERAPEUTICS, INC.** [US/US]: 2001 Aliceanna Street, Baltimore, MD 21231-2001 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LIU, Xuan** [CN/US]; 14231 Dav Road, Rockville, MD 20850 (US). **CHENG, Linzhao** [CN/US]; 5346 Woodlot Road, Columbia, MD 21044 (US).

(74) Agents: **GRANT, Alan, J.** et al.; Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road, Roseland, NJ 07068 (US).

(81) Designated States *(national)*: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(48) Date of publication of this corrected version:
11 October 2001

(15) Information about Correction:
see PCT Gazette No. 41/2001 of 11 October 2001, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/63382 A1

(54) Title: GENES AND EXPRESSION PRODUCTS FROM HEMATOPOIETIC CELLS

(57) Abstract: A human hematopoietic stem/precursor cell (hHSPC) polypeptide (called C17 polypeptide), as well as DNA (and RNA) encoding such polypeptide, are disclosed. Also disclosed are methods for utilizing the polynucleotides and polypeptides disclosed herein, including as markers for chromosomal mapping, DNA fingerprinting and the possible role played by genetic mutations in the disease process, and for the generation of polyclonal sera or monoclonal antibodies specific for said polypeptides. Also disclosed is a method for increasing the rate of multiplication of hMSCs utilizing the polypeptide of the invention.

GENES AND EXPRESSION PRODUCTS FROM HEMATOPOIETIC CELLS

This application claims priority of U.S. Provisional Application 60/129,643, filed April 15, 1999, the disclosure of which is hereby incorporated in its entirety.

BACKGROUND OF THE INVENTION

This invention relates to newly identified polynucleotide sequences corresponding to transcription products of human genes, and to complete gene sequences associated therewith and to gene expression products thereof and to uses for the foregoing, especially where these involve hematopoiesis and the bone marrow microenvironment. More specifically, the invention disclosed herein relates to a novel gene that is expressed in $CD34^+$ hematopoietic stem and/or progenitor cells (HSPCs) but not in $CD34^-$ hematopoietic cells.

It is well settled that circulating blood cells are products of the terminal differentiation of a number of determined precursor cells. During fetal life, hematopoiesis occurs throughout the reticuloendothelial system whereas in the adult the terminal differentiation of precursor cells (for example, precursors of white cells, red cells and platelets) occurs only in the bone marrow, especially that of the axial skeleton. Marrow films and biopsy specimens have contributed much information about the condition of the hematopoietic process in the living organism, especially humans.

Researchers have defined a pluripotent stem cell, which can give rise to any and all types of blood cells. In particular, this pluripotent stem cell, found in the marrow, will differentiate along one of two well defined pathways. Thus, the stem cell will differentiate into either a myeloid stem cell, ultimately giving rise to all of the final differentiated blood cells except B and T lymphocytes, or will differentiate into a lymphoid stem cell, which itself will eventually differentiate into either plasma cells or T lymphocytes (T cells). Current research into the nature of hematological disease as well as into the formation of differentiated blood cells has tended to focus on these progenitor cells, thereby giving rise to what has been called the progenitor basis of hematopoiesis.

It would be very difficult to follow the development of various types of blood cells within the marrow compartment if it were not for the discovery of the presence on the surfaces of such cells of various cell surface antigens, proteins specific to a particular type of cell or a particular group of cell types. As cells differentiate, the pattern of cell surface antigens on their surfaces changes as various genes within the cells are either turned off or turned on with each successive generation.

Most such markers have been classified using the established "CD" nomenclature, especially useful when describing hematopoietic cells. The term "CD" has been understood as describing either "cluster designation" or "cluster of differentiation" and refers to a molecule recognized by a "cluster" of monoclonal antibodies useful in identifying the stage of differentiation of the cells and thus to distinguish one class of hematological cells from another, including cells operating at different stages of the hematopoietic process.

These CD proteins are routinely used to determine the identity of various types of cells and to follow the progress of hematopoiesis from precursor stem cells to final differentiated progeny.

For example, CD34 is a protein of about 105 to 120 kD in size and is present on hematopoietic stem cells.

Human hematopoietic stem/progenitor cells (HSPC) are enriched in a rare population of bone marrow (BM) mononuclear cells (MNC) that bear the CD34 surface antigen ($CD34^+$). In addition to BM, $CD34^+$ HSPC are also found in neonatal cord blood (CB), and are enriched in peripheral blood after mobilization by cytokines and chemotherapy (Krause et al., *Blood*, **87**, 1 (1996). $CD34^+$ cells account for ~1-4% of MNC in BM and ~0.5% in CB and mobilized peripheral blood (mPB). Purified $CD34^+$ cells can engraft in bone marrow and generate blood/lymphoid cells for years in patients after transplantation. In addition, single $CD34^+$ cells can form colonies of hematopoietic cells in culture. In contrast, the counterpart mononuclear cells that lack CD34 expression ($CD34^-$) are largely mature hematopoietic cells of various differentiated lineages, and have lost the ability to form colonies in culture (Krause et al., 1996).

The invention herein discloses diagnostic and therapeutic applications of a novel secreted protein and its encoding gene, the latter gene being specifically expressed in $CD34^+$ hematopoietic stem/progenitor cells.

BRIEF SUMMARY OF THE INVENTION

Human hematopoietic stem/progenitor cells (hHSPCs) are part of a relatively rare population of mononuclear cells present in bone marrow and in blood from the umbilical cord and which express the CD34 cell surface antigen (such cells being denoted $CD34^+$). However, the CD34 gene is not exclusively expressed in HSPCs (for example, endothelial cells also express high levels of CD34).

In accordance with the present invention, blood cells obtained from umbilical cord were found to express a novel protein if the cells possessed the CD34 cell surface antigen but was not expressed if CD34 was not detected on the cells.

It is an object of the present invention to use these cells to prepare cDNA clones of novel genes, and expression products thereof, including nucleic acids, isolated sequences, and fragments thereof, and use these for the determination and preparation of the expression products of these nucleic acids and sequences, including fragments thereof.

It is still another object of the present invention to provide protein expression products of the cDNAs, such as mRNAs, as well as other nucleic acids, especially isolated nucleic acids, nucleotide sequences of such nucleic acids, and fragments thereof, identified according to the present invention that are expressed in hematopoietic stem/progenitor cells.

It is another object of the present invention to provide polypeptide expression products of the nucleotide sequences disclosed herein, which polypeptides act as growth factors for mesenchymal stem cells, which cells are capable of differentiating into most types of connective tissue cell, such stem cells being useful, for example, in replacement therapies and the like.

It is a further object of the present invention to use the cDNAs so produced, and fragments thereof, as well as their expression products, as chromosomal markers for determining the location of such genes, including any alleles thereof, within the genome.

It is yet another object of the present invention to provide DNA sequences for use in human "fingerprinting" whereby different individuals can be distinguished based on the sequences of the genes identified as wholly, or partly identical, to those disclosed herein.

It is still another object of the present invention to provide polynucleotide sequences corresponding to the genes coding for polypeptides as disclosed herein whereby such sequences can be compared with those found in similar chromosomal locations in mammals, especially humans, where such mammal is afflicted with a disease, thereby detecting the presence of mutations in said genes, said mutations possibly leading to such diseases.

It is a still further object of the present invention to provide genetically engineered cells, and vectors, containing one or more copies of the nucleic acids, or DNAs, or genes, or nucleotide sequences according to the present invention, capable of expressing the peptides, or polypeptides, or proteins, according to the present invention, for rapid cloning thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the nucleotide sequence for the novel gene disclosed according to the present invention (SEQ ID NO: 1) which contains the open reading frame (SEQ ID NO: 2) corresponding to the deduced protein of Figure 2 and the two relevant dpnII restriction sites useful in cloning the gene and which provide the fragment utilized. Figure 1B is a continuation of Figure 1A.

Figure 2 shows the deduced amino acid sequence (SEQ ID NO:3) for the open reading frame of the sequence disclosed in Figure 1 and thus for the C17 protein.

Figure 3 shows the effect of C17 protein on proliferation of mesenchymal stem cells in serum-free culture. The C17 protein allows for proliferation rates equivalent to that supported by serum-containing medium.

DETAILED DESCRIPTION OF THE INVENTION

One aspect of the present invention is directed to nucleic acids and isolated DNA sequences and molecules, and fragments thereof (and corresponding isolated RNA sequences, and fragments thereof) showing sequence homology with, or capable of hybridizing to, the DNA sequence identified in Figure 1 (SEQ ID NO: 1). The present invention is also directed to fragments or portions of such sequences which contain at least 15 bases, preferably at least 30 bases, more preferably at least 50 bases and most preferably at least 80 bases, and to those sequences which are at least 60%, preferably at least 80%, and most preferably at least 95% identical thereto, and to DNA (or RNA) sequences encoding the same polypeptide as the sequence of Figure 1, including fragments and portions thereof and, when derived from natural sources, includes alleles thereof.

In accordance with the present invention, the term "percent identity" or "percent identical," when referring to a sequence, means that a sequence is compared to a claimed or described sequence after alignment of the sequence to be compared (the "Compared Sequence") with the described or claimed sequence (the "Reference Sequence"). The Percent Identity is then determined according to the following formula:

Percent Identity = 100 [1-(C/R)]

wherein C is the number of differences between the Reference Sequence and the Compared Sequence over the length of alignment between the Reference Sequence and the Compared Sequence wherein (i) each base or amino acid in the Reference Sequence that does not have a corresponding aligned base or amino acid in the Compared Sequence and (ii) each gap in the Reference Sequence and (iii) each aligned base or amino acid in the Reference Sequence that is different from an aligned base or amino acid in the Compared Sequence, constitutes a difference; and R is the number of bases or amino acids in the Reference Sequence over the length of the alignment with the Compared Sequence with any gap created in the Reference Sequence also being counted as a base or amino acid.

If an alignment exists between the Compared Sequence and the Reference Sequence in which the percent identity as calculated above is about equal to or greater than a specified minimum Percent Identity then the Compared Sequence has the specified minimum percent identity to the Reference Sequence even though alignments may exist in which the hereinabove calculated Percent Identity is less than the specified Percent Identity.

A further aspect of the present invention is directed to a DNA sequence (as well as the corresponding RNA sequence) which is or contains a DNA sequence identical to one contained in Figure 1 (SEQ ID NO: 1). A DNA sequence according to the present invention is hybridizable under stringent conditions with a DNA sequence identified in Figure 1 and set forth in the Sequence Listing (Seq. ID No. 1). As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 97% identity between the sequences.

Yet another aspect of the present invention is directed to an isolated DNA (or RNA) sequence or molecule comprising at least the coding region of a human gene (or a DNA sequence encoding the same polypeptide as such coding region), in particular an expressed human gene, which human gene comprises a DNA sequence homologous with, or contributing to, the sequence depicted in Figure 1 (SEQ ID NO: 1), or one at least 90%, preferably at least 95%, and most preferably at least 98%, identical thereto, as well as fragments or portions of the coding region which encode a polypeptide having a similar function to the polypeptide encoded by said coding region. Thus, the isolated DNA (or RNA) sequence can include only the coding region of the expressed gene (or fragment or portion thereof as hereinabove indicated) or can further include all or a portion of the non-coding DNA (or RNA) of the expressed human gene.

In general, sequences homologous with and contributing to the sequence shown in Figure 1 (SEQ ID NO: 1), or one at least 90%, preferably at least 95%, and most preferably at least 98% identical or homologous thereto, are from the coding region of a human gene.

The present invention also relates to vectors or plasmids which include such DNA (or RNA) sequences, as well as the use of the DNA (or RNA) sequences.

The sequence depicted in Figure 1 (SEQ ID NO: 1), is hybridizable with actual DNA and RNA sequences as derived from different human tissues. The distribution of this sequence in various human tissues was determined from database matchings for other human sequences.

The polynucleotides of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic

DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the mature polypeptide may be identical to the coding sequence shown in Figure 1 or may be a different coding sequence, which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature polypeptide as the DNA of Figure 1 (SEQ ID NO: 1).

The polynucleotide which codes for the polypeptide of Figure 2 (SEQ ID NO: 3) may include, but is not limited to: only the coding sequence for the mature polypeptide; the coding sequence for the mature polypeptide and additional coding sequence such as a leader or secretory sequence, a proprotein sequence and a membrane anchor; the coding sequence for the mature polypeptide (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature polypeptide.

Thus, the term "polynucleotide" as used for the present invention encompasses a polynucleotide which includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequences.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the polypeptide having the amino acid sequence of Figure 2 (SEQ ID NO: 3). Variants of the polynucleotide may be naturally occurring allelic variants of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the nucleic acids, or polynucleotides, according to the present invention may have coding sequences which are naturally occurring allelic variants of the coding sequence shown in Figure 1. As known in the

art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

The present invention also includes polynucleotides, wherein the coding sequence for the mature polypeptide may be fused in the same reading frame to a polynucleotide sequence which aids in expression and secretion of a polypeptide from a host cell, for example, a leader sequence which functions as a secretory sequence for controlling transport of a polypeptide from the cell and a transmembrane anchor which facilitates attachment of the polypeptide to a cellular membrane. The polypeptide having a leader sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature polypeptide. The polynucleotides may also encode for a proprotein which is the mature protein plus additional 5' amino acid residues. A mature protein having a prosequence is a proprotein and is often an inactive form of the protein. Once the prosequence is cleaved an active mature protein remains.

Thus, for example, a polynucleotide according to the present invention may code for a mature protein, for a protein having a prosequence, for a protein having a transmembrane anchor or for a polypeptide having a prosequence, a presequence (leader sequence) and a transmembrane anchor.

The polynucleotides of the present invention may also have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptide of the present invention. The marker sequence may be a hexa-histidine tag supplied by a pQE-9 vector to provide for purification of the mature polypeptide fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used.

The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell, 37:767 (1984)).

Fragments of the full length polynucleotide of the present invention may be used as hybridization probes for a cDNA library to isolate the full length cDNA and to isolate other cDNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type preferably have at least 15 bases, may have at least 30 bases and even 50 or more bases. The probe may also be used to identify a cDNA clone corresponding to a full-length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons, and introns. An example of a screen comprises isolating the coding region of the gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary to that of the gene of the present invention are used to screen a library of human cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

A polynucleotide according to the present invention may have at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. Such polynucleotides may be employed as probes for the polynucleotide of Figure 1, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

The polynucleotides according to the present invention may also occur in the form of mixtures of polynucleotides hybridizable to some extent with the sequence of Figure 1 (SEQ ID NO: 1), including any and all fragments thereof, and which polynucleotide mixtures may be composed of any number of such polynucleotides, or fragments thereof, including

mixtures having at least 10, perhaps at least 30 such sequences, or fragments thereof.

Because coding regions comprise only a small portion of the human genome, identification and mapping of transcribed regions and coding regions of chromosomes is of significant interest. There is a corresponding need for reagents for identifying and marking coding regions and transcribed regions of chromosomes. Furthermore, such human sequences are valuable for chromosome mapping, human identification, identification of tissue type and origin, forensic identification, and locating disease-associated genes (i.e., genes that are associated with an inherited human disease, whether through mutation, deletion, or faulty gene expression) on the chromosome.

Various aspects of the present invention include each of the individual sequences, corresponding partial and complete cDNAs, genomic DNA, mRNA, antisense strands, PCR primers, coding regions, and constructs. Expression vectors and polypeptide expression products, are also within the scope of the present invention, along with antibodies, especially monoclonal antibodies, to such expression products.

As used herein and except as noted otherwise, all terms are defined as given below.

In accordance with the present invention, the term "gene" or "cistron" means the segment of DNA (or DNA segment) involved in producing a polypeptide chain; it includes regions preceding and following the coding region (5'-and 3'- untranslated regions, or UTRs, also called leader and trailer sequences, regions, or segments) as well as intervening sequences (introns) between individual coding segments (exons), which intronic regions are typically removed during processing of post-

transcriptional RNA to form the final translatable mRNA product. Of course, by their nature, cDNAs contain no intronic sequences.

In accordance with the present invention, the term "DNA segment" refers to a DNA polymer, in the form of a separate fragment or as a component of a larger DNA construct, which has been derived from DNA isolated at least once in substantially pure form, i.e., free of contaminating endogenous materials and in a quantity or concentration enabling identification, manipulation, and recovery of the segment and its component nucleotide sequences by standard biochemical methods, for example, using a cloning vector. Such segments are provided in the form of an open reading frame uninterrupted by internal nontranslated sequences, or introns, which are typically present in eukaryotic genes. Sequences of non-translated DNA may be present downstream from the open reading frame, where the same do not interfere with manipulation or expression of the coding regions.

The nucleic acids and polypeptide expression products disclosed according to the present invention, as well as expression vectors containing such nucleic acids, may be in "enriched form." As used herein, the term "enriched" means that the concentration of the material is at least about 2, 5, 10, 100, or 1000 times its natural concentration (for example), advantageously 0.01%, by weight, preferably at least about 0.1% by weight. Enriched preparations of about 0.5%, 1%, 5%, 10%, and 20% by weight are also contemplated. The sequences, constructs, vectors, clones, and other materials comprising the present invention can advantageously be in enriched or isolated form. For example, removal, via the differential display techniques described herein, of clones corresponding to ribosomal RNA and "housekeeping" genes and clones without human cDNA inserts results in a library that is "enriched" in the desired clones.

The DNA and RNA sequences, and polypeptides, disclosed in accordance with the present invention will commonly be in isolated form. The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or DNA present in a living animal is not isolated, but the same polynucleotide or DNA, separated from some or all of the coexisting materials in the natural system, is isolated. Such DNA could be part of a vector and/or such polynucleotide could be part of a composition, and still be isolated in that such vector or polynucleotide is not part of its natural environment.

The DNA and RNA sequences, or polypeptides, disclosed in accordance with the present invention may also be in "purified" form. The term "purified" does not require absolute purity; rather, it is intended as a relative definition, and can include preparations that are highly purified or preparations that are only partially purified, as those terms are understood by those of skill in the relevant art. Individual clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The cDNA clones are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). By conversion of mRNA into a cDNA library, pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from RNA and subsequently isolating individual clones from that library results in an approximately 10^6 fold purification of the native message. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. Furthermore, claimed polynucleotide which has a purity of preferably 0.001%, or at least 0.01% or 0.1%; and even desirably 1% by weight or greater is expressly contemplated.

The term "coding region" refers to that portion of a human gene which either naturally or normally codes for the expression product of that gene in its natural genomic environment, i.e., the region coding *in vivo* for the native expression product of the gene. The coding region can be from a normal, mutated or altered gene, or can even be from a DNA sequence, or gene, wholly synthesized in the laboratory using methods well known to those of skill in the art of DNA synthesis.

In accordance with the present invention, the term "nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the proteins provided by this invention are assembled from cDNA fragments and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

The term "expression product" means that polypeptide or protein that is the natural transcription product of the gene and any nucleic acid sequence coding equivalents resulting from genetic code degeneracy and thus coding for the same amino acid(s).

The term "fragment," when referring to a coding sequence, means a portion of DNA comprising less than the complete human coding region whose expression product retains essentially the same biological function or activity as the expression product of the complete coding region.

The term "primer" means a short nucleic acid sequence that is paired with one strand of DNA and provides a free 3'OH end at which a DNA polymerase starts synthesis of a deoxyribonucleotide chain.

The term "promoter" means a region of DNA involved in binding of RNA polymerase to initiate transcription.

The term "open reading frame (ORF)" means a series of triplets coding for amino acids without any termination codons and is a sequence (potentially) translatable into protein.

The term "exon" means any segment of an interrupted gene that is represented in the mature RNA product.

As used herein, reference to a DNA sequence includes both single stranded and double stranded DNA. Thus, the specific sequence, unless the context indicates otherwise, refers to the single strand DNA of such sequence, the duplex of such sequence with its complement (double stranded DNA) and the complement of such sequence.

In accordance with the present invention, the overall approach to identification of cDNAs involved with the mesenchymal differentiation process in hMSCs involved measurement of gene expression during osteogenic differentiation of the cells as grown in culture. Cells were harvested and the total RNA content thereof was recovered. Next, using various primer combinations, reverse transcriptase and polymerase chain reaction procedures were used to produce and amplify the corresponding cDNAs, which were then screened to find regulated DNA sequences that were subsequently purified and cloned. These clones were then sequenced and used to determine a consensus sequence (one based upon the most commonly occurring bases at each nucleotide position in a sequence after the contributing sequences are aligned by residue position). The resulting sequences were then subjected to computer database searches for novelty, and any homology with known sequences, using, for example, the BLAST program and the GenBank database.

In accordance with the foregoing, a cDNA library was generated and used to identify the sequence of Figure 1 (SEQ ID NO: 1). Probes based on these cDNAs were used to identify the relevant transcripts, using Northern Blotting Analysis methods well known in the art.

The nucleotide sequence disclosed according to the present invention, as contained in Figure 1 (SEQ ID NO: 1 and 2) was found to be expressed in CD34 bearing cells of cord blood as well as of bone marrow, with the full transcript being about 1.1 kb (as determined by Northern Blot Hybridization Analysis). The sequence contained an open reading frame coding for a polypeptide of 136 amino acids (the latter showing no significant homology to any of the known proteins in GenBank and was therefore considered to be novel). Hydropathy analysis of the deduced peptide sequence (shown in Figure 2, SEQ ID NO: 3) indicates a signal peptide of 19 amino acids at the N-terminus of the protein, suggesting that it is secreted by $CD34^+$ cells. This secretion characteristic was confirmed by tagging the cDNA with a histidine-6 tag (the latter allowing ready purification by Nickel affinity chromatography) and with epitopes of the human *c-myc* gene, at locations corresponding to the C-terminus of the expressed protein. When the recombinant protein was expressed by human 293 cells, it was found to be secreted into the medium.

Each of the DNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. The sequences can be used as diagnostic probes for the presence of a specific mRNA in a particular cell type as well as in genetic linkage analysis (polymorphisms). Further, the sequences can be used as probes for locating gene regions associated with genetic disease.

The nucleotide and gene sequences of the present invention are also valuable for chromosome identification. Each sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. The mapping of the polynucleotides to specific chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease, such as diseases affecting bone formation or skeletal abnormalities.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-30 bp) from the sequences disclosed herein. Computer analysis of these sequences is used to rapidly select primers that do not span more than one exon in the corresponding genomic DNA, which would otherwise complicate the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the sequences or subsequences disclosed herein will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more clones can be assigned per day using a single thermal cycler, as is well known in the art. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map a sequence, or part of a sequence, to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence in situ hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 500 or 600 bases; however, clones larger than 2,000 bp have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. FISH requires use of the clone from which the sequence was derived, and the longer the better. For example, 2,000 bp is good, 4,000 is better, but more than 4,000 is probably not necessary to get good results a reasonable percentage of the time. For a review of this technique, see Verma et al., Human Chromosomes: a Manual of Basic Techniques. Pergamon Press, New York (1988).

Reagents for chromosome mapping can be used individually (to mark a single chromosome or a single site on that chromosome) or as panels of reagents (for marking multiple sites and/or multiple chromosomes). Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library)). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically close genes).

Next, it is necessary to determine if there are differences in the cDNA or genomic sequence between affected and unaffected

individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb.)

Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that cDNA sequence. Ultimately, complete sequencing of genes from several individuals is required to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

In addition to the foregoing, the sequences of the invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al, Nucl. Acids Res., 6:3073 (1979); Cooney et al, Science, 241:456 (1988) ; and Dervan et al, Science, 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem., 56:560 (1991) ; Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of

the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

The present invention is also a useful tool in gene therapy, which requires isolation of the disease-associated gene in question as a prerequisite to the insertion of a normal gene into an organism to correct a genetic defect. The high specificity of the cDNA probes according to this invention have promise of targeting such gene locations in a highly accurate manner.

The sequences of the present invention, as broadly defined, and including subsequences and fragments thereof, are also useful for identification of individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP.

However, RFLP is a pattern based technique, which does not require the DNA sequence of the individual to be sequenced. Portions of the sequences of the present invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can also be used to prepare PCR primers for amplifying and isolating such selected DNA. One can, for example, take part of the sequence of the invention and prepare two PCR primers from the 5' and 3' ends of the sequence, or fragment of the sequence. These are used to

amplify an individual's DNA, corresponding to the sequence. The amplified DNA is sequenced.

Panels of corresponding DNA sequences from individuals, made this way, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences, due to allelic differences. The sequences of the present invention can be used to particular advantage to obtain such identification sequences from individuals and from tissue. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the fragments or complete coding sequences comprising a part of the present invention can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals.

If a panel of reagents from the sequences according to the present invention is used to generate a unique ID database for an individual, those same reagents can later be used to identify tissue from that individual. Positive identification of that individual, living or dead can be made from extremely small tissue samples.

Another use for DNA-based identification techniques is in forensic biology. PCR technology can be used to amplify DNA sequences taken from very small biological samples. In one prior art technique, gene sequences are amplified at specific loci known to contain a large number of allelic variations, for example the DQα class II HLA gene (Erlich, H., PCR Technology, Freeman and Co. (1992)). Once this specific area of the genome is amplified, it is digested with one or more restriction

enzymes to yield an identifying set of bands on a Southern blot probed with DNA corresponding to the DQα class II HLA gene.

The sequences of the present invention can be used to provide polynucleotide reagents specifically targeted to additional loci in the human genome, and can enhance the reliability of DNA-based forensic identifications. Those sequences targeted to noncoding regions are particularly appropriate. As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Reagents for obtaining such sequence information are within the scope of the present invention. Such reagents can comprise complete genes, parts of genes or corresponding coding regions, or fragments of at least 15 bp, preferably at least 18 bp.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar manner, these reagents can be used to screen tissue cultures for contamination.

Sequences that match perfectly to several different genes can be detected by hybridizing to chromosomes: if many chromosomal loci are observed, the sequence (or a close variant) is in more than one gene. This problem can be circumvented by using the 3'-untranslated part of the cDNA alone as a probe for the chromosomal location or for the full-length cDNA or gene. The 3'-untranslated region is more likely to be unique within gene families, since there is no evolutionary pressure to conserve a coding function of this region of the mRNA.

The cDNA libraries disclosed according to the present invention ideally use directional cloning methods so that either the 5' end of the cDNA (likely to contain coding sequence) or the 3' end (likely to be a non-coding sequence) can be selectively obtained.

Using the sequence information provided herein, the polynucleotides of the present invention can be derived from natural sources or synthesized using known methods. The sequences falling within the scope of the present invention are not limited to the specific sequences described, but include human allelic and species variations thereof and portions thereof. In addition, the invention includes the entire coding sequence associated with the specific polynucleotide sequence of bases described in the Sequence Listing, as well as portions of the entire coding sequence. Allelic variations can be routinely determined by comparison of one sequence with a sequence from another individual of the same species. Furthermore, to accommodate codon variability, the invention includes sequences coding for the same amino acid sequences as do the specific sequences disclosed herein. In other words, in a coding region, substitution of one codon for another which encodes the same amino acid is expressly contemplated. (Coding regions can be determined through routine sequence analysis.)

In a cDNA library there are many species of mRNA represented. Each cDNA clone can be interesting in its own right, but must be isolated from the library before further experimentation can be completed. In order to sequence any specific cDNA, it must be removed and separated (i.e. isolated and purified) from all the other sequences. This can be accomplished by many techniques known to those of skill in the art. These procedures normally involve identification of a bacterial colony containing the cDNA of interest and further amplification of that bacteria. Once a cDNA is separated from the mixed clone library, it can

be used as a template for further procedures such as nucleotide sequencing.

The present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

Thus, the present invention is not restricted to such constructs or sequences alone but also includes expression vehicles, which may include plasmids, viruses, or any other expression vectors, including cells and liposomes, containing any of the nucleic acids, nucleotide sequences, DNAs, RNAs, or fragments thereof, as disclosed according to the present invention. Furthermore, this will be true regardless of whether such sequences are coding sequences or non-coding sequences and whether such coding sequences code for all or part of the expression products as disclosed herein, so long as such expression products, or fragments thereof, exhibit some utility in keeping with the invention disclosed herein. Thus, while the present invention includes an isolated DNA sequence, or nucleic acid, that expresses a human protein when in a suitable expression system, for example, a cell-free, or *in vitro*, expression system, such system may also be contained

in, or part of, a suitable expression vehicle, or vector, be that a cell, a plasmid, a virus, or other operative expression vector.

Such expression systems, especially where part of an expression vehicle, will commonly require some promoter region that may include a promoter different from that normally associated *in vivo* with the genes coding for the gene expression products and proteins disclosed according to the present invention. Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described construct(s). The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a procaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, 1986)) .

The constructs in host cells can be used in a conventional manner to produce the gene product coded by the recombinant sequence. Alternatively, the encoded polypeptide, once the sequence is known from the cDNAs, or from isolation of the pure product, can be synthetically produced by conventional methods of peptide synthesis, either manual or automated.

Thus, in accordance with the present invention, once the coding sequence is known, or the gene is cloned which encodes the polypeptide, conventional techniques in molecular biology can be used to obtain the polypeptide. More generally, the present invention includes all polypeptides coded for by any and each of the DNA or RNA sequences disclosed herein, including fragments of said polypeptides, as well as derivatives and functional analogs thereof.

At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. (Fragments are useful, for example, in generating antibodies against the native polypeptide.)

Alternatively, the DNA encoding the desired polypeptide can be inserted into a host organism and expressed. The organism can be a bacterium, yeast, cell line, or multicellular plant or animal. The literature is replete with examples of suitable host organisms and expression techniques. For example, polynucleotide (DNA or mRNA) can be injected directly into muscle tissue of mammals, where it is expressed. This methodology can be used to deliver the polypeptide to the animal, or to generate an immune response against a foreign polypeptide. Wolff, et al., Science, 247:1465 (1990); Felgner, et al., Nature, 349:351 (1991). Alternatively, the coding sequence, together with appropriate regulatory regions (i.e., a construct), can be inserted into a vector, which is then used to transfect a cell. The cell (which may or may not be part of a larger organism) then expresses the polypeptide.

The present invention further relates to a polypeptide which has the amino acid sequence of Figure 2, (SEQ ID NO: 3)as well as fragments, analogs and derivatives of such polypeptide.

The terms "fragment," "derivative" and "analog," when referring to the polypeptide of Figure 2 (SEQ ID NO: 3), means a polypeptide which retains essentially the same biological function or activity as said polypeptide. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide. Such fragments, derivatives and analogs must have sufficient similarity to the polypeptide of Figure 2 (SEQ ID NO: 3) so that activity of the native polypeptide is retained.

The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a synthetic polypeptide, preferably a recombinant polypeptide.

"Recombinant," as used herein, means that a protein is derived from recombinant (e.g., microbial or mammalian) expression systems. "Microbial" refers to recombinant proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Protein expressed in most bacterial cultures, e.g., E. coli, will be free of glycosylation modifications; protein expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

The fragment, derivative or analog of the polypeptide of Figure 2 (SEQ ID NO: 3) may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the

additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the abilities of those skilled in the art in view of the teachings herein.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity. When applied to polypeptides, the term "isolated" has its already stated meaning.

The polypeptides of the present invention include the polypeptide of Figure 2 (in particular the mature polypeptide) as well as polypeptides which have at least 90% identity to the polypeptide of Figure 2 (SEQ ID NO: 3), or which have, at least 95% identity to the polypeptide of Figure 2 (SEQ ID NO: 3) and still more preferably at least 98% identity to the polypeptide of Figure 2 (SEQ ID NO: 3) and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

In accordance with the present invention, the polypeptide disclosed in Figure 2 has growth stimulating activity when present in an *in vitro* growth medium containing human mesenchymal stem cells. Thus, such stem cells, in the presence of the polypeptide disclosed herein, are induced to replicate at a faster rate (as shown in Figure 3). Thus,

recombinant C17 protein, expressed by 293 cells, was affinity purified and added to human MSC (hMSC) cultures. The hMSCs, maintained in serum-free conditions, typically exhibit little basal proliferative activity. Here, dos titrations of fetal calf serum (FBS) were used as a positive control. Recombinant C17 protein stimulated hMSC growth by about 10 fold compared to serum-free media, and at levels equivalent to 10% fetal calf serum. The C17 polypeptide would be present in the medium at a concentration of at least 1 picogram (pg) per ml of medium.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector, either of which may be in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia,

adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

In accordance with the present invention, an appropriate DNA sequence or segment may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into the appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (for example, a promoter sequence) to direct mRNA synthesis. Representative examples of such promoters are: LTR or SV40 promoter, the *E. coli. lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in E. coli.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as E. coli, Streptomyces, Salmonella typhimurium; fungal cells, such as yeast; insect cells such as Drosophila S2 and Spodoptera Sf9; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

"Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous protein upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, (Cold Spring Harbor, N.Y., 1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the polypeptides according to the present invention by higher eukarotes can be increased by insertion of an enhancer sequence into the vector. Such enhancers have been known for some time and are usually cis-acting elements of DNA, usually anywhere from 10 to 300 bp that act on a promoter to increase transcription. Common examples include the SV40 enhancer, the cytomegalovirus early promoter enhancer, the polyoma enhancer and the enhancers found in adenovirus.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired

characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli, Bacillus subtilis, Salmonella typhimurium* and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is derepressed by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant protein produced in bacterial culture is conveniently isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

The protein, its fragments or other derivatives, or analogs thereof, or cells expressing them, can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal, monoclonal, chimeric, single chain, Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of polyclonal antibodies.

Antibodies generated against the polypeptide corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptide into an animal or by administering the polypeptide to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies binding the whole native polypeptide. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide. Moreover, a panel of such antibodies, specific to a large number of polypeptides, can be used to identify and differentiate such tissue.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention.

The antibodies can be used in methods relating to the localization and activity of the protein sequences of the invention, e.g., for imaging these proteins, measuring levels thereof in appropriate physiological samples and the like.

Of course, knowing the sequence of the C17 protein of Figure 2 will permit those skilled in the art to readily locate appropriate

receptors on the surfaces of the mesenchymal stem cells, as well as other cell types, and thereby confer the ability to regulate growth of such cells.

In addition, since the present invention also encompasses sequences homologous to the disclosed nucleotide and polypeptide sequences, it will of course be possible to derive structurally similar analogs containing similar functional domains, including small molecules that can mimic the functions of the C17 protein without themselves being proteinaceous in structure. Thus, small organic molecules may easily be developed by molecular modeling, using computer programs and algorithms, or by combinatorial methods, to mimic the domains of the C17 protein disclosed herein. Such mimicing structures are also considered to be encompassed by the disclosure of the present invention. Such chemicals can be readily synthesized and added to cell growth media, thereby stimulating the relevant receptors and enhancing the rate of cell growth. Such methods of enhancing cell growth are likewise deemed to be within the bounds of the invention disclosed herein.

Such growth effects can easily be used to locate cell-growth stimulating receptors on the surfaces of cells. Here, cells can be grown in a suitable medium to which has been added an appropriate amount of a labeled C17 protein, or homolog thereof, or small chemical analog thereof, and then determining if said homolog, or analog, can stimulate the growth of the cells. If so, the analog can also be introduced in a suitably labeled form, typically chemically labeled by the usual means well known to chemists, such labeling including both radiolabeled and nonradiolabeled methods, and then allowed to remain in the medium for various periods of time to allow for possible binding to a surface receptor on the surface of the cells so as to locate such receptors. This can then be followed by use of common isolation techniques to permit isolating, identification and characterization of the receptors, be they surface receptors or otherwise.

In so doing, the growth-stimulating receptors of various cell types can be determined.

In carrying out the procedures of the present invention it is of course to be understood that reference to particular buffers, media, reagents, cells, culture conditions and the like are not intended to be limiting, but are to be read so as to include all related materials that one of ordinary skill in the art would recognize as being of interest or value in the particular context in which that discussion is presented. For example, it is often possible to substitute one buffer system or culture medium for another and still achieve similar, if not identical, results. Those of skill in the art will have sufficient knowledge of such systems and methodologies so as to be able, without undue experimentation, to make such substitutions as will optimally serve their purposes in using the methods and procedures disclosed herein.

Specific embodiments of the invention will now be further described in more detail in the following non-limiting examples and it will be appreciated that additional and different embodiments of the teachings of the present invention will doubtless suggest themselves to those of skill in the art and such other embodiments are considered to have been inferred from the disclosure herein.

METHODS AND MATERIALS

Cord blood cells and CD34$^+$ cell isolation

Frozen cord blood cells from full-term newborns were purchased from the Cord Blood Bank at University of Arizona. Mononuclear cells (MNC) from 3-4 units of CB were pooled and labeled with an anti-CD34 antibody (clone QBEND/10) provided in the CD34 Progenitor Cell Isolation

Kit (Miltenyi Biotec, Auburn, CA). Up to 2 billion MNC were passed through an LS$^+$ column assembled in the VarioMACS system (Miltenyi Biotec). The CD34$^+$ cells that were labeled with magnetic beads and retained in the column were isolated by eluting cells from the column after removal from the magnet. To ensure the elimination of CD34$^-$ cells in the flow-through (FT) fraction, these FT cells were passed through a second column as before. The FT fraction after this double depletion was used as the CD34$^-$ cell population. CD34$^+$ cells isolated from the first and second column were pooled. The content of CD34$^+$ cells of each population was monitored using the fluorescence-activated cell sorting (FACS) staining (see below). The majority of cells were immediately lysed with TRIzol reagent (Gibco/BRL, Gaithersburg, MD).

Other human cells

CD34$^+$ cells from bone marrow of healthy donors were isolated similarly by the PureCell Company (San Mateo, CA), according to federal and state regulations. We also used human CD34$^+$ cells from mPB from healthy volunteers. Five days after consecutive G-CSF treatment, leukopheresed blood cells were obtained. CD34$^+$ and CD34$^-$ cells were isolated similarly, using the Isolex-300 system (Nexell/Baxter, Irvine, CA). Bone marrow-derived mesenchymal stem cells (MSC) were isolated and expanded in culture as described in the literature (Pittenger et al., *Science*, **284**, 143 (1999)).

FACS analysis of CD34$^-$ cells

CB cells before and after cell isolation were labeled with a R-Phycoerythrin (R-PE)-conjugated CD34 antibody (Clone HPCA-2, Becton Dickinson Immunology Systems [BDIS], San Jose, CA). HPCA-2 recognizes a different CD34 epitope from that recognized by QBEND/10, which is used to purify the cells. Antibody-labeled cells were analyzed with a BDIS FACS Calibur or Vantage instrument equipped with an ion

Argon laser tuned to 488 nm. Specific CD34 staining of individual MNC was recorded in the FL2 channel (for R-PE). Non-specific staining (background) was 0.1%. The content of $CD34^+$ cells before cell isolation was 0.4 +/- 0.1% (n = 3), consistent with previous publications (Cairo and Wagner, *Blood*, **90**:4665-4678 (1997)). Typically the percentages of $CD34^+$ cells in the $CD34^+$ preparations were 85.0%, and in $CD34^-$ fractions they were approximately 0.1% (the background level). Among the 30 CB samples processed, a few cell preparations did not meet these criteria and were discarded.

Complementary DNA (cDNA) Synthesis

Total RNA was isolated using TRIzol Reagent (Gibco/BRL). Two hundred micrograms of total RNA were isolated from 40 million $CD34^+$ cells pooled from several preparations. Approximately 2 micrograms of poly A^+ RNA were purified using the mRNA Isolation Midi Kit (Qiagen, Valencia, CA). Double-stranded cDNA was prepared using the cDNA Synthesis System containing the Superscript II reverse transcriptase and random hexamers (Gibco/BRL). RNA samples derived from $CD34^+$ and $CD34^-$ cells were always processed in parallel.

Generation and Analysis of RDA Gene Fragments

Representational difference analysis (RDA) amplicon preparation and subtractive hybridization was done as described in the literature (Lisitsyn et al, *Science*, **259**:946-951 (1993); Hubank and Schatz, *Nucleic Acids Research*, **22**:5640-5648 (1994)), except that shorter PCR cycles (95°C, 30 sec, 72°C, 2 min) were used for preparation of amplicons (before subtraction) and difference products (after subtraction). After three rounds of subtraction, distinct bands were apparent in an agarose gel. The third (and final) difference products (DP3) were digested with DpnII (to remove adapter and generate GATC overhangs), and then cloned into a BamHI-digested pUC18 vector. More than 500 clones were

obtained after we transformed the DH5α strain of *E. coli* with a small aliquot of the ligated DNA. Initially 55 individual clones were randomly picked. The inserts of individual clones were PCR amplified and sequenced. The sequences were searched first against the GenBank non-redundant (NR) database using the BLASTN and BLASTX algorithms (Altschul et al., *Nucleic Acids Res.*, **25**:3389-3402 (1997)); http://www.ncbi.nlm.nih.gov/BLAST/). Those with no significant matches were considered to be novel. The novel sequences were then searched against the GenBank EST database (dbEST) using the BLASTN algorithm.

Oligonucleotide primers for RT-PCR

The primers for CD34 cDNA amplification (298 bp) are as follows:

CD34-5': CTGTGTCTCAACATGGCA-3' (SEQ ID NO: 4)
CD34-3': GCCTTGATGTCACTTAGG-5' (SEQ ID NO: 5)

The primers for C17 cDNA amplification (286 bp) are:

C17-5': GATCACCCGCGACTTCAACC (SEQ ID NO: 6)
C17-3': TGGCAGGACCGTAGTCACTG (SEQ ID NO: 7)

The primers for beta-2-microglobulin (β2M, as a control) cDNA amplification (270 bp) are:

β2M-5': TCTGGCCTTGAGGCTATCCAGCGT (SEQ ID NO: 8)
β2M-3': GTGGTTCACACGGCAGGCATACTC (SEQ ID NO: 9)

Plasmids containing C17 cDNA

In addition to our RDA clones containing the C17 cDNA fragment (290 bp), we purchased 5 plasmids containing human ESTs (Table 1) from Research Genetics, Inc. (Huntsville, Alabama), a distributor of EST clones for the International Molecular Analysis of Gene Expression (IMAGE) Consortium. The IMAGE clone 786066 contains the longest insert (~1 kb) and a region identical to C17, and was used as the source of the C17 coding sequence for subsequent analyses.

Recombinant gene expression in human cells

The complete coding region of C17 from the IMAGE clone 786066 was amplified by PCR and cloned in-frame into the mammalian expression vector pCDNA3.1/myc-HisB (Invitrogen, Carlsbad, CA), at the EcoRI and BamHI sites. The resulting plasmid is named pCMV.C17/myc/his. The recombinant C17 protein expressed from this construct is tagged with a human c-myc epitope and six histidine residues (His6) at the C-terminus (in italics below).

As a result, 31 amino acids

(VDPSSVPSFL*EQKLISEEDL*NSAVD*HHHHHH*) (SEQ ID NO: 10)

were added to the C-terminus of C17.

Human 293T-derived BOSC23 cells were transfected with the vector by calcium phosphate precipitation (Cheng et al., *Nature Biotech.*, 14:606 (1996); Cheng et al, *Gene Ther.*, 4:1013 (1997)). Forty-eight hours after the transfection, the cells and the conditioned media were collected. The cells were scraped from the culture dishes in the presence of a protein inhibitor cocktail (Complete™; Roche Biochemicals). The cells were lysed in a buffer containing 150mM NaCl, 20mM Tris-HCl (pH7.4),

10% glycerol, 1% NP40, 10mM EDTA, 2mM $NaVO_3$, 100mM NaF, and the Complete™. The lysates were cleared by centrifugation at 14,000 rpm for 30 min at 4°C. The cell extracts and the conditioned culture media were denatured in the sample buffer under reducing conditions, and electrophoresed on a 4-20% polyacrylamide gel in SDS-Tris/Glycine buffer. The production of rC17 was monitored by Western blot with antibodies against the c-myc or His6 epitope (from Invitrogen).

Chromosomal Mapping

The Stanford G3 human-hamster radiation hybrid (RH) panel was purchased from Research Genetics. Two pairs of PCR primers were designed based on the 5'-untranslated region of C17 cDNA:

TTTGATTTTCATCACCTTTC (DEQ ID NO: 11)
and CTGGTTTAATGGAGTAATGG (SEQ ID NO: 12)

GTTAGATACACAGCATGTTGA (SEQ ID NO: 13)
and GACAGTGAAGAAAGTCTGTG (SEQ ID NO: 14)

Each pair specifically amplified a ~200 bp DNA fragment with a genomic DNA template from human but not from hamster. PCR reactions were performed using 25 ng (nanograms) of a DNA template under the following condition: 94°C, 20 sec; 55°C, 20 sec; 72°C, 30 sec for 30 cycles. Both sets of primers gave an identical result. The result of the PCR reactions was tabulated (using 1 as positive and 0 as negative), and submitted to the Stanford Human Genome Center RHserver (http://www-shgc.stanford.edu/). LOD scores higher than 6 are considered significant.

Table 1. Human EST entries related to the C17 RDA fragment

EST Id	Derived tissue	Score (bits)	E value	IMAGE clone #	Sequenced insert (bp)	Estimated insert size (bp)
AA448744	9 wk whole embryo	573	e-132	786066	431	ND
T81361	Fetal liver/spleen	297	5e-79	110792	397	830
T82005	Fetal liver/spleen	295	2e-78	110293	436	520
AA460463	9 wk whole embryo	143	2e-32	796569	464	ND
AA461037	9 wk whole embryo	143	2e-32	796773	549	ND

The C17 gene fragment (290 bp) was searched against dbEST using the BLASTN algorithm. See http://ncbi.nlm.nih.gov/blast/ for more details of the cDNA libraries used, score and E value. ND: not determined by the depositors, who partially sequenced the inserts either from 5' or 3' ends. At the time of the search (July 1998), a total of 2,072,964 EST (human and non-human) entries had been deposited.

EXAMPLE 1

Preferential gene expression of C17 in $CD34^-$ hematopoietic cells was verified as follows. RT-PCR (reverse transcriptase-polymerase chain reaction) was performed using total RNA from $CD34^+$ or $CD34^-$ cells and two C17-specific primers (based on the sequence of a 290 bp C17 RDA fragment). The C17 gene expression was readily detected in CB $CD34^-$ cells but was undetectable in the $CD34^-$ cell population. A similar RT-PCR result was obtained with the cells from mPB as well as with the cells from bone marrow. Therefore, the C17 gene expression is restricted to the $CD34^-$ cell population isolated from CB, BM and mPB, three sources known to contain HSPC.

C17 gene expression in untreated and cultured $CD34^-$ cells was analyzed by Northern blot. BM $CD34^-$ cells were cultured under two culture conditions with different cytokines. Under the first condition, BM $CD34^+$ cells were treated with TPO, SCF and Flt3/Flk2 ligand (FL), a combination which is known to favor the maintenance of stem cells and expansion of progenitor cells (Luens et al., *Blood*, **91**:1206-1215 (1998); Kaushansky, *Blood*, **92**:1-3 (1998)). Under the second condition, cells were treated with five hematopoietic colony-stimulating factors (IL-3, IL-6, G-CSF, GM-CSF and EPO). They are known to expand committed progenitor cells and stimulate cell differentiation, resulting in HSPC differentiation into myeloid/erythroid lineages and $CD34^+$ cell reduction (Luens et al., 1998). C17 gene expression in cultured and untreated $CD34^-$ cells was analyzed by Northern blot using the C17 RDA fragment (290 bp) as the probe. A single prominent band of ~1.0 kb was observed in untreated BM $CD34^+$ cells as well as in cultured cells, which expressed C17 gene at various levels. After culture for 7 to 15 days, the C17 mRNA level was elevated under condition #1 while it was slightly reduced under condition #2.

In normal tissues, using Clontech's multiple tissue blots containing purified polyA+ RNA, C17 expression was detected in human bone marrow and very weakly in lymph nodes, but undetectable in spleen, thymus, fetal liver, and PBL by Northern hybridization. In the same hybridization, C17 was undetectable in the another blot containing polyA+ RNA from several human cancer cell lines: HeLa S3 (cervical carcinoma), A549 (lung carcinoma), G-361 (melanoma), SW480 (colorectal adenocarcinoma), HL-60 (promyelocytic leukemia), K-562 (chronic myelogenous leukemia), Molt-4 (lymphoblastic leukemia), and Raji (Burkitt's lymphoma). In addition, Northern blot and RT-PCR analyses showed that the C17 gene transcript was absent in mesenchymal stem cells, which are believed to reside in close proximity to HSPC in bone

marrow (Pittenger et al., 1999). Thus, C17 expression is regulated by hematopoietic cytokines.

EXAMPLE 2

A full-length DNA sequence of the C17 cDNA was obtained by purchasing five plasmid clones containing C17-related EST sequences (Table 1). The insert of these plasmids has been partially sequences from either the 5' or 3' end by the IMAGE Consortium members. The size of inserts in these plasmids was determined and sequenced from both ends. The insert in IMAGE clone 786066 is the longest (~1 kb), and includes all the sequences from the other four plasmids and the C17 RDA fragment. The insert sequence of IMAGE clone 786066 was used for subsequent analyses. A putative mRNA polyadenylation signal, AATAAA, is found near the 3' end of the C17 cDNA (see, for example, SEQ ID NO: 1 at residues 979-984). Based on the transcript size (~ 1 kb) shown in Northern blots and sequences from multiple ESTs, the IMAGE clone 786066 contains a near-full length cDNA for the C17 gene. The C17 cDNA contains an open-reading-frame of 408 nucleotides, encoding a protein of 136 amino acids (SEQ ID NO: 3). The presence of a Kozak sequence immediately around the first ATG suggests that it is a favorable translation start (Kozak, *J. Cell Biol.*, **115**:887-903 (1991)). A hydrophobicity analysis of the deduced peptide sequence shows a putative signal peptide at the N-terminus. Moreover, a defined signal peptide analysis revealed that the secretory peptide cleavage site is between the 19th and 20th amino acids thereof (Nielsen et al., *Protein Engineering*, **10**:1-6 (1997)). There are no other hydrophobic transmembrane or GPI-anchoring signal domains in the rest of the sequence, indicating that C17 is a secreted protein. Secondary-structure analysis predicts that the C17 peptide contains 4 alpha helices, a characteristic of hematopoietic cytokines and interleukins (Bazan,

Immunology Today, 11(10):350-354 (1990); Wells and de Vos, *Ann. Rev. Biochem.*, 65:609-634 (1996)). The GenBank accession number for C17 is AF193766.

EXAMPLE 3

The C17 protein was further characterized by cloning the C17 cDNA coding region into a mammalian expression vector to make pCMV.C17/myc/his. The recombinant C17 protein was tagged with both the 9E10 *c-myc* epitope and six histidine residues (His6) at the C-terminus, thereby facilitating detection and purification of rC17. Human 293T cells were transfected with the vector to allow expression of the tagged C17 gene. Forty-eight hours after transfection, both the cell extracts and the conditioned media (supernatants) from transfected cells were analyzed by Western blot using antibodies against either of the two tags. Anti-myc antibody recognized specific proteins in the cell extract and supernatant unique to the C17-transfected cells. In the cell extract, a major 19 kD protein band is specifically recognized, which is consistent with the predicted size of 19 kD for the unprocessed, tagged C17 protein (167 amino acids total, including the signal peptide). In the supernatant collected from C17-transfected cells, a single protein band was detected, indicating that the C17 protein was indeed secreted. The apparent molecular weight (approximately 26 kD) was larger than the predicted size of 17.2 kD (the processed C17 protein without the 19 amino acid signal peptide), suggesting that the secreted protein was modified during or after secretion. Based on the amino acid sequence, there are no potential N-glycosylation sites in C17 or in the epitope tags. Digestion with a panel of glycosidases also failed to shift the protein migration in SDS-PAGE.

EXAMPLE 4

rC17 was produced in large quantity by cloning C17 into a prokaryotic expression vector, pBAD/gIII (from Invitrogen). In this expression vector, the putative C17 signal sequence was removed and the rest of the coding sequence was ligated in frame with a bacterial leader sequence. This allows the recombinant protein to be secreted into the pericytoplasmic space. The C17 protein expressed by this vector is also tagged with the c-myc and His6 epitopes. Upon induction by arabinose, a protein of 19 kD (as predicted for the full-length rC17) was induced to express at a high level by 0.002% or higher arabinose concentration.

EXAMPLE 5

The radiation hybrid (RH) technique was used to map the location of the C17 gene in the human genome. A panel of G3 human-hamster hybrid chromosomal DNA samples was used as templates for PCR amplification with primers specific to human C17 gene. The primers can only amplify a 200 bp fragment if human (but not hamster) genomic DNA is present as a template. PCR reactions with some G3 RH DNA templates generated the predicted DNA fragment while the others failed. The resulting data were used to determine its chromosomal localization, based on the Stanford Human Genome Center database and algorithm (http://www-shgc.stanford.edu/). The unique pattern mapped the C17 gene to a single locus in human chromosome 4p, between D4S412 and D4S1601 (http://www.ncbi.nlm.nih.gov/genemap98/map.cgi?MAP=G3 &BIN=130&MARK=SHGC-33462). This result is confirmed by the mapping of its related ESTs (Hs.13872 in the Unigene database) performed by others. This region co-localizes with human chromosome 4p15-16 in cytogenetic mapping. Some other genes associated with

hematopoiesis are also localized in this region. These include CD38 (between D4S412 and D4S1601, as C17/Hs.13872) and the AC133 antigen (located around D4S1601 to D4S1608). The latter is a recently discovered cell surface protein and is preferentially expressed in $CD34^-$ HSPC (Yin et al., *Blood*, **90**:5002-5012 (1997); Miraglia et al., *Blood*, **90**:5013-5021 (1997)).

WHAT IS CLAIMED IS:

1. An isolated nucleic acid comprising a polynucleotide that is at least 90% identical to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Figure 2.

2. An isolated nucleic acid comprising a polynucleotide that is at least 95% identical to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Figure 2.

3. An isolated nucleic acid comprising a polynucleotide that is at least 98% identical to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Figure 2.

4. An isolated nucleic acid comprising RNA complementary to any of the DNA sequences or fragments of claim 1, 2 or 3.

5. An isolated nucleic acid comprising a DNA sequence identical to the DNA sequence of Figure 1.

6. An isolated nucleic acid comprising RNA complementary to the DNA sequence of Claim 5.

7. An isolated nucleic acid comprising at least the polypeptide coding region of a human gene, said human gene containing a DNA sequence according to Claim 1.

8. An isolated nucleic acid comprising at least the polypeptide coding region of a human gene which contains the DNA sequence of Claim 5.

9. The isolated nucleic acid of claim 8 which expresses a human protein when in a suitable expression system.

10. An expression vehicle comprising the DNA sequence of claim 3.

11. An expression vehicle comprising the DNA sequence of claim 5.

12. An expression vehicle comprising the DNA sequence of claim 7.

13. An expression vehicle comprising the DNA sequence of claim 9.

14. A polypeptide coded for by the DNA sequence of claim 7 and active fragments, derivatives and functional analogs thereof.

15. A polypeptide coded for by the DNA sequence of claim 8 and active fragments, derivatives and functional analogs thereof.

16. A polypeptide comprising the amino acid sequence of Figure 2 (SEQ ID NO: 3).

17. A genetically engineered cell having inserted into the genome thereof the DNA of Claim 7.

18. A process for producing cells for expressing a polypeptide using genetically engineering cells claim 27.

19. An isolated DNA sequence comprising a fragment of DNA of Figure 1 (SEQ ID NO: 1), wherein said fragment comprises at least 15 sequential bases of said sequence.

20. An isolated DNA sequence comprising a fragment of DNA of Figure 1 (SEQ ID NO: 1), wherein said fragment comprises at least 30 sequential bases of said sequence.

21. An isolated DNA sequence comprising a fragment of DNA of Figure 1 (SEQ ID NO: 1), wherein said fragment comprises at least 50 sequential bases of said sequence.

22. An isolated DNA sequence comprising a fragment of DNA of Figure 1 (SEQ ID NO: 1), wherein said fragment comprises at least 80 sequential bases of said sequence.

23. An antiserum prepared by immunizing a mammal with a polypeptide according to claims 14 or 15.

24. A monoclonal antibody against the polypeptide according to claim 14.

25. A monoclonal antibody against the polypeptide according to claim 15.

26. A method for increasing the rate of multiplication of human mesenchymal stem cells *in vitro* comprising adding an effective amount of the polypeptide of claim 16 to the extracellular growth medium in which such cells are suspended.

27. The method of claim 26 wherein the polypeptide is present in said growth medium at a concentration of at least 1 pg/ml.

28. A chemical compound having a structure similar to a domain of the C17 polypeptide of Figure 2.

29. A method of determining the presence of growth-stimulating receptors on the surface of a cell comprising the use of the chemically labeled C17 polypeptide of Figure 2, or analogs thereof, to detect the presence of such receptor(s) on the surface of such cell by incubating such analog with cells in a suitable medium, determining if there is growth stimulation by the presence of such analog, detecting the analog bound to the cell surface and isolating the receptor to which such analog is bound.

FIGURE 1A

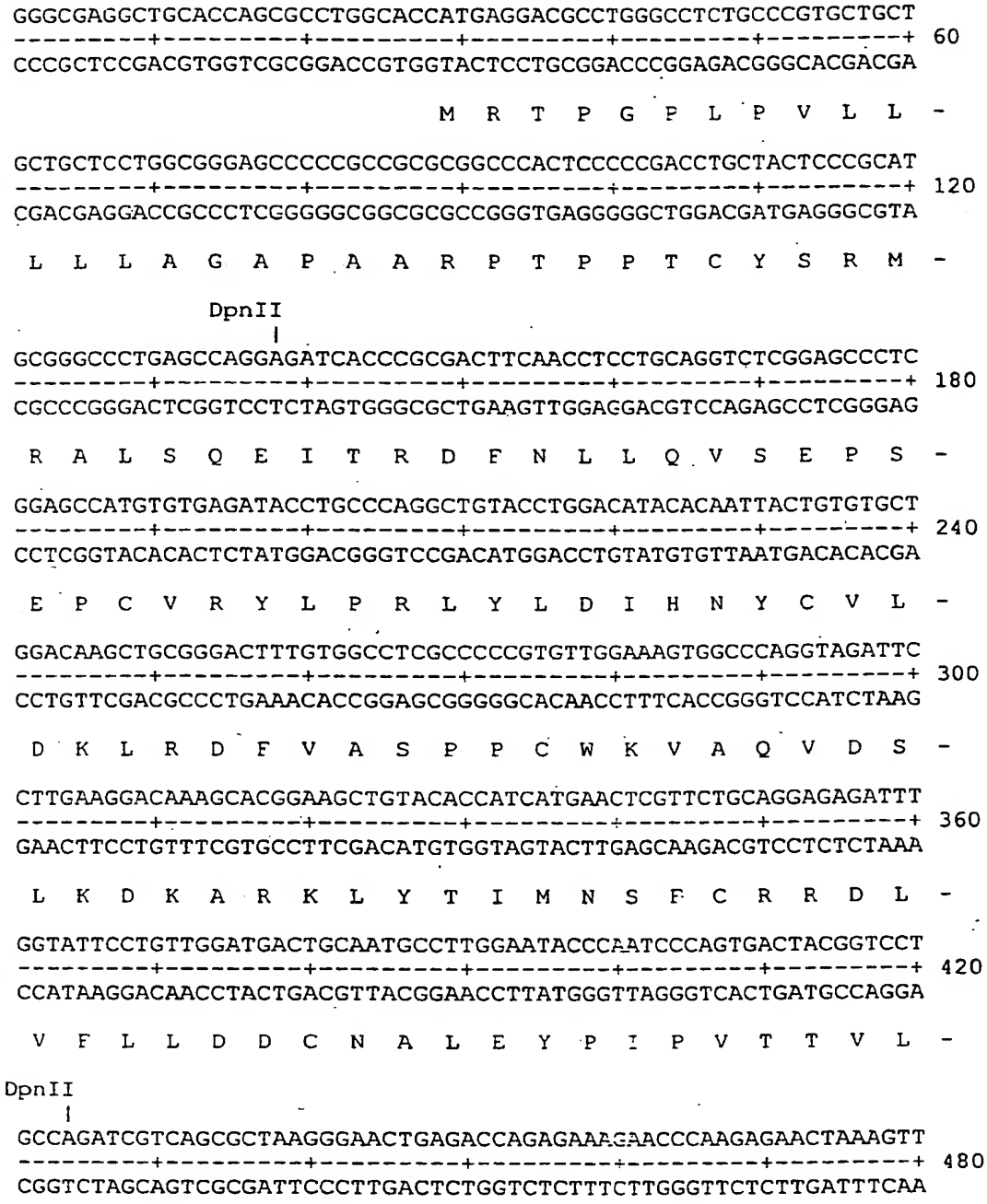

GGGCGAGGCTGCACCAGCGCCTGGCACCATGAGGACGCCTGGGCCTCTGCCCGTGCTGCT
CCCGCTCCGACGTGGTCGCGGACCGTGGTACTCCTGCGGACCCGGAGACGGGCACGACGA
M R T P G P L P V L L -
GCTGCTCCTGGCGGGAGCCCCCGCCGCGCGGCCCACTCCCCCGACCTGCTACTCCCGCAT
CGACGAGGACCGCCCTCGGGGGCGGCGCGCCGGGTGAGGGGGCTGGACGATGAGGGCGTA
L L L A G A P A A R P T P P T C Y S R M -
DpnII
GCGGGCCCTGAGCCAGGAGATCACCCGCGACTTCAACCTCCTGCAGGTCTCGGAGCCCTC
CGCCCGGGACTCGGTCCTCTAGTGGGCGCTGAAGTTGGAGGACGTCCAGAGCCTCGGGAG
R A L S Q E I T R D F N L L Q V S E P S -
GGAGCCATGTGTGAGATACCTGCCCAGGCTGTACCTGGACATACACAATTACTGTGTGCT
CCTCGGTACACACTCTATGGACGGGTCCGACATGGACCTGTATGTGTTAATGACACACGA
E P C V R Y L P R L Y L D I H N Y C V L -
GGACAAGCTGCGGGACTTTGTGGCCTCGCCCCCGTGTTGGAAAGTGGCCCAGGTAGATTC
CCTGTTCGACGCCCTGAAACACCGGAGCGGGGGCACAACCTTTCACCGGGTCCATCTAAG
D K L R D F V A S P P C W K V A Q V D S -
CTTGAAGGACAAAGCACGGAAGCTGTACACCATCATGAACTCGTTCTGCAGGAGAGATTT
GAACTTCCTGTTTCGTGCCTTCGACATGTGGTAGTACTTGAGCAAGACGTCCTCTCTAAA
L K D K A R K L Y T I M N S F C R R D L -
GGTATTCCTGTTGGATGACTGCAATGCCTTGGAATACCCAATCCCAGTGACTACGGTCCT
CCATAAGGACAACCTACTGACGTTACGGAACCTTATGGGTTAGGGTCACTGATGCCAGGA
V F L L D D C N A L E Y P I P V T T V L -
DpnII
GCCAGATCGTCAGCGCTAAGGGAACTGAGACCAGAGAAAGAACCCAAGAGAACTAAAGTT
CGGTCTAGCAGTCGCGATTCCCTTGACTCTGGTCTCTTTCTTGGGTTCTCTTGATTTCAA

FIGURE 1B

```
     P  D  R  Q  R  *
    ATGTCAGCTACCCAGACTTAATGGGCCAGAGCCATGACCCTCACAGGTCTTGTGTTAGTT
481 ---------+---------+---------+---------+---------+---------+ 540
    TACAGTCGATGGGTCTGAATTACCCGGTCTCGGTACTGGGAGTGTCCAGAACACAATCAA

    GTATCTGAAACTGTTATGTATCTCTCTACCTTCTGGAAAACAGGGCTGGTATTCCTACCC
541 ---------+---------+---------+---------+---------+---------+ 600
    CATAGACTTTGACAATACATAGAGAGATGGAAGACCTTTTGTCCCGACCATAAGGATGGG

    AGGAACCTCCTTTGAGCATAGAGTTAGCAACCATGCTTCTCATTCCCTTGACTCATGTCT
601 ---------+---------+---------+---------+---------+---------+ 660
    TCCTTGGAGGAAACTCGTATCTCAATCGTTGGTACGAAGAGTAAGGGAACTGAGTACAGA

    TGCCAGGATGGTTAGATACACAGCATGTTGATTTGGTCACTAAAAAGAAGAAAAGGACTA
661 ---------+---------+---------+---------+---------+---------+ 720
    ACGGTCCTACCAATCTATGTGTCGTACAACTAAACCAGTGATTTTTCTTCTTTTCCTGAT

    ACAAGCTTCACTTTTATGAACAACTATTTTGAGAACATGCACAATAGTATGTTTTTATTA
721 ---------+---------+---------+---------+---------+---------+ 780
    TGTTCGAAGTGAAAATACTTGTTGATAAAACTCTTGTACGTGTTATCATACAAAAATAAT

    CTGGTTTAATGGAGTAATGGTACTTTTATTCTTTCTTGATAGAAACCTGCTTACATTTAA
781 ---------+---------+---------+---------+---------+---------+ 840
    GACCAAATTACCTCATTACCATGAAAATAAGAAAGAACTATCTTTGGACGAATGTAAATT

    CCAAGCTTCTATTATGCCTTTTTCTAACACAGACTTTCTTCACTGTCTTTCATTTAAAAA
841 ---------+---------+---------+---------+---------+---------+ 900
    GGTTCGAAGATAATACGGAAAAAGATTGTGTCTGAAAGAAGTGACAGAAAGTAAATTTTT

    GAAATTAATGCTCTTAAGATATATATTTTACGTAGTGCTGACAGGACCCACTCTTTCATT
901 ---------+---------+---------+---------+---------+---------+ 960
    CTTTAATTACGAGAATTCTATATATAAAATGCATCACGACTGTCCTGGGTGAGAAAGTAA

    GAAAGGTGATGAAAATCAAATAAAGAATCTCTTCACATG
961 ---------+---------+---------+--------- 999
    CTTTCCACTACTTTTAGTTTATTTCTTAGAGAAGTGTAC
```

FIGURE 2

Met Arg Thr Pro Gly Pro Leu Pro Val Leu Leu Leu Leu Leu Ala Gly

Ala Pro Ala Ala Arg Pro Thr Pro Pro Thr Cys Tyr Ser Arg Met Arg

Ala Leu Ser Gln Glu Ile Thr Arg Asp Phe Asn Leu Leu Gln Val Ser

Glu Pro Ser Glu Pro Cys Val Arg Tyr Leu Pro Arg Leu Tyr Leu Asp

Ile His Asn Tyr Cys Val Leu Asp Lys Leu Arg Asp Phe Val Ala Ser

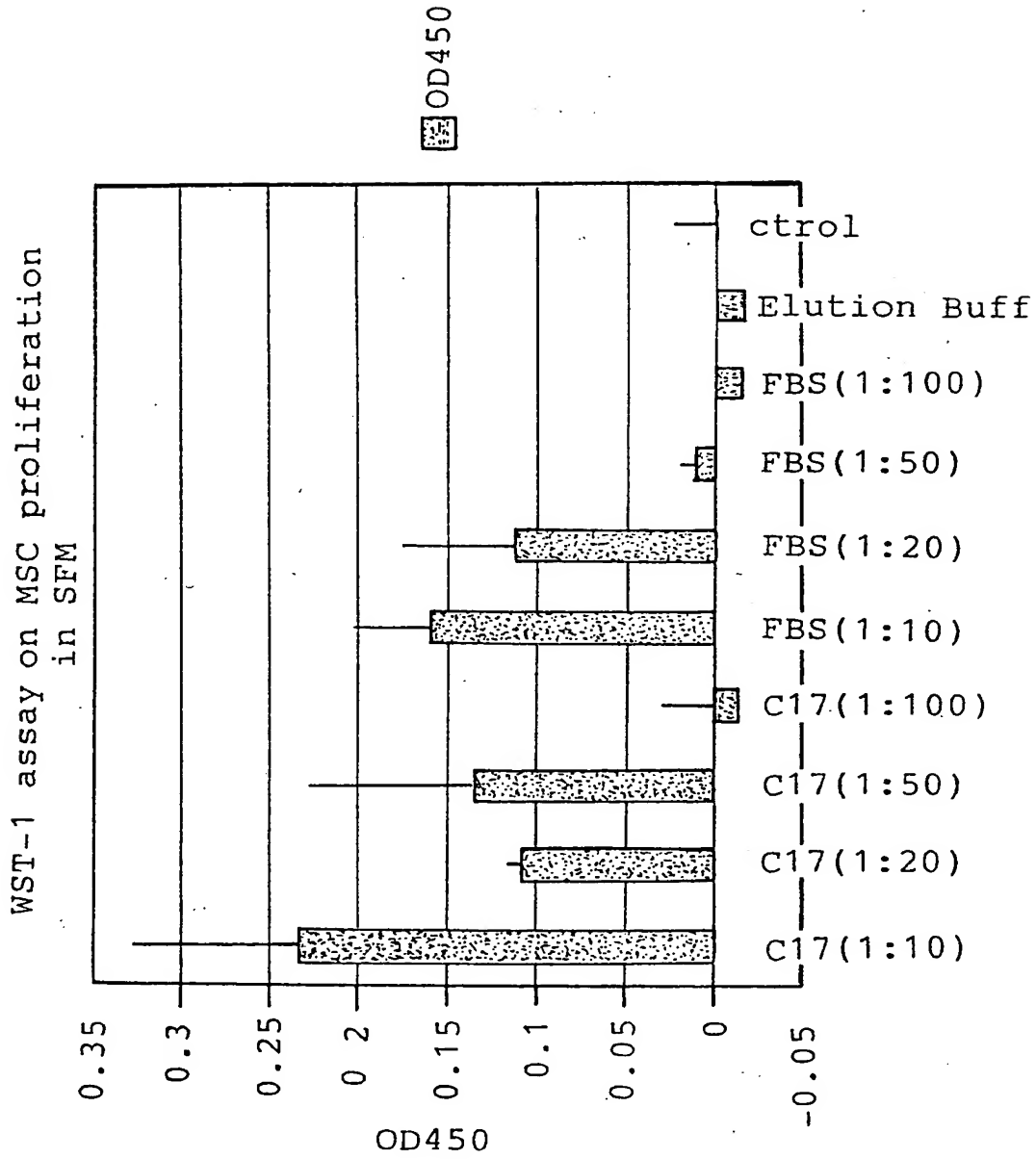
Pro Pro Cys Trp Lys Val Ala Gln Val Asp Ser Leu Lys Asp Lys Ala

Arg Lys Leu Tyr Thr Ile Met Asn Ser Phe Cys Arg Arg Asp Leu Val

Phe Leu Leu Asp Asp Cys Asn Ala Leu Gly Tyr Pro Ile Pro Val Thr

Thr Val Leu Pro Asp Arg Glu Arg

FIG. 3


OD450
WST-1 assay on MSC proliferation in SFM
ctrol
Elution Buff
FBS(1:100)
FBS(1:50)
FBS(1:20)
FBS(1:10)
C17(1:100)
C17(1:50)
C17(1:20)
C17(1:10)
0.35
0.3
0.25
0 2
0.15
0.1
0.05
0
-0.05
OD450

SEQUENCE LISTING

<110> Liu, Xuan
Cheng, Linzhou

<120> Novel Genes and Expression Products from Hematopoietic Cells

<130> 640100-364

<140>
<141>

<150> U.S. 60/129,463
<151> 1999-04-15

<160> 14

<170> PatentIn Ver. 2.1

<210> 1
<211> 999
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Consensus sequence for C17 protein

<400> 1
gggcgaggct gcaccagcgc ctggcaccat gaggacgcct gggcctctgc ccgtgctgct 60
gctgctcctg gcgggagccc ccgccgcgcg gcccactccc ccgacctgct actcccgcat 120
gcgggccctg agccaggaga tcacccgcga cttcaacctc ctgcaggtct cggagccctc 180
ggagccatgt gtgagatacc tgcccaggct gtacctggac atacacaatt actgtgtgct 240
ggacaagctg cgggactttg tggcctcgcc cccgtgttgg aaagtggccc aggtagattc 300
cttgaaggac aaagcacgga agctgtacac catcatgaac tcgttctgca ggagagattt 360
ggtattcctg ttggatgact gcaatgcctt ggaataccca atcccagtga ctacggtcct 420
gccagatcgt cagcgctaag ggaactgaga ccagagaaag aacccaagag aactaaagtt 480
atgtcagcta cccagactta atgggccaga gccatgaccc tcacaggtct tgtgttagtt 540
gtatctgaaa ctgttatgta tctctctacc ttctggaaaa cagggctggt attcctaccc 600
aggaacctcc tttgagcata gagttagcaa ccatgcttct cattcccttg actcatgtct 660
tgccaggatg gttagataca cagcatgttg atttggtcac taaaaagaag aaaaggacta 720
acaagcttca cttttatgaa caactatttt gagaacatgc acaatagtat gtttttatta 780
ctggtttaat ggagtaatgg tacttttatt ctttcttgat agaaacctgc ttacatttaa 840
ccaagcttct attatgcctt tttctaacac agactttctt cactgtcttt catttaaaaa 900
gaaattaatg ctcttaagat atatatttta cgtagtgctg acaggaccca ctctttcatt 960
gaaaggtgat gaaaatcaaa taaagaatct cttcacatg 999

<210> 2
<211> 411
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:cDNA sequence
for C17 protein

<400> 2
atgaggacgc ctgggcctct gcccgtgctg ctgctgctcc tggcgggagc ccccgccgcg 60
cggcccactc ccccgacctg ctactcccgc atgcgggccc tgagccagga gatcacccgc 120
gacttcaacc tcctgcaggt ctcggagccc tcggagccat gtgtgagata cctgcccagg 180
ctgtacctgg acatacacaa ttactgtgtg ctggacaagc tgcgggactt tgtggcctcg 240
cccccgtgtt ggaaagtggc ccaggtagat tccttgaagg acaaagcacg gaagctgtac 300
accatcatga actcgttctg caggagagat ttggtattcc tgttggatga ctgcaatgcc 360
ttggaatacc caatcccagt gactacggtc ctgccagatc gtcagcgcta a 411

<210> 3
<211> 136
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Amino acid
sequence derived from cDNA

<400> 3
Met Arg Thr Pro Gly Pro Leu Pro Val Leu Leu Leu Leu Leu Ala Gly
1 5 10 15

Ala Pro Ala Ala Arg Pro Thr Pro Pro Thr Cys Tyr Ser Arg Met Arg
20 25 30

Ala Leu Ser Gln Glu Ile Thr Arg Asp Phe Asn Leu Leu Gln Val Ser
35 40 45

Glu Pro Ser Glu Pro Cys Val Arg Tyr Leu Pro Arg Leu Tyr Leu Asp
50 55 60

Ile His Asn Tyr Cys Val Leu Asp Lys Leu Arg Asp Phe Val Ala Ser
65 70 75 80

Pro Pro Cys Trp Lys Val Ala Gln Val Asp Ser Leu Lys Asp Lys Ala
85 90 95

Arg Lys Leu Tyr Thr Ile Met Asn Ser Phe Cys Arg Arg Asp Leu Val
100 105 110

Phe Leu Leu Asp Asp Cys Asn Ala Leu Glu Tyr Pro Ile Pro Val Thr
115 120 125

Thr Val Leu Pro Asp Arg Gln Arg
130 135

<210> 4
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PCR forward primer for CD34 cDNA

<400> 4
ctgtgtctca acatggca 18

<210> 5
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PCR reverse primer for CD34 cDNA

<400> 5
gccttgatgt cacttagg 18

<210> 6
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PCR forward primer for C17 cDNA

<400> 6
gatcacccgc gacttcaacc 20

<210> 7
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PCR reverse primer for C17 cDNA

<400> 7
tggcaggacc gtagtcactg 20

<210> 8
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PCR forward primer for beta-2-microglobulin cDNA

<400> 8
tctggccttg aggctatcca gcgt 24

<210> 9
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PCR reverse primer for beta-2-microglobulin cDNA

<400> 9
gtggttcaca cggcaggcat actc 24

<210> 10
<211> 31
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Human c-myc epitope used to tag recombinant C17 protein

<400> 10
Val Asp Pro Ser Ser Val Pro Ser Phe Leu Glu Gln Lys Leu Ile Ser
1 5 10 15

Glu Glu Asp Leu Asn Ser Ala Val Asp His His His His His His
20 25 30

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PCR primer based on 5'-untranslated region of C17 cDNA

<400> 11
tttgattttc atcacctttc 20

<210> 12
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PCR primer based on 5'-untranslated region of C17 cDNA

<400> 12
ctggtttaat ggagtaatgg 20

<210> 13
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PCR primer based on 5'-untranslated region of C17 cDNA

<400> 13
gttagataca cagcatgttg a 21

<210> 14
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PCR primer
based on 5'-untranslated region of C17 cDNA

<400> 14
gacagtgaag aaagtctgtg 20

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/09904

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/52 G01N33/50 C07K16/24

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, STRAND

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 54963 A (FERRIE ANN M;HUMAN GENOME SCIENCES INC ; GREENE JOHN M (US); YOUNG) 10 December 1998 (1998-12-10) SEQ ID NOs 72,295 page 328 -page 329 page 516 -page 517 ---	1-5, 7-22,28
X	HILLIER L ET AL: "Homo sapiens cDNA clone" EMEST DATABASE ENTRY HS1248747, ACCESSION NUMBER AA448744, 10 June 1997 (1997-06-10), XP002143937 sequence --- -/--	19-22

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 July 2000	03.08.00

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Espen, J

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HILLIER L ET AL: "Homo sapiens cDNA clone" EMEST DATABASE ENTRY HS1151114, ACCESSION NUMBER AA233071, 6 March 1997 (1997-03-06), XP002143938 sequence -----	19-22

INTER[redacted]IONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 00/09904

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9854963 A	10-12-1998	AU 7812098 A	21-12-1998
		AU 7811998 A	21-12-1998
		AU 6552198 A	29-09-1998
		EP 0973892 A	26-01-2000
		WO 9840483 A	17-09-1998